DOCUMENT RESUME

ED 423 244 TM 028 973

TITLE	Student Testing: Issues Related to Voluntary National Mathematics and Reading Tests. Report to the Honorable William F. Goodling, Chairman, Committee on Education and the Workforce, House of Representatives, and the Honorable John Ashcroft, U.S. Senate.
INSTITUTION	General Accounting Office, Washington, DC. Health, Education, and Human Services Div.
REPORT NO	GAO/HEHS-98-163
PUB DATE	1998-06-00
NOTE	45p.
PUB TYPE	Reports - Evaluative (142)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Agency Cooperation; Costs; *Educational Testing; Elementary Education; Elementary School Students; Grade 4; Grade 8; Mathematics Achievement; *National Competency Tests; Reading Achievement; *State Programs; Student Characteristics; Student Evaluation; *Test Construction; Test Results; *Testing Programs
IDENTIFIERS	Department of Education; National Assessment Governing Board; National Assessment of Educational Progress; *Voluntary Participation

ABSTRACT

Information was sought about three main issues related to voluntary national tests for fourth-grade reading and eighth-grade mathematics: (1) the relationship between the National Assessment Governing Board and the U.S. Department of Education; (2) costs of developing the national voluntary tests and procedures for hiring contractors; and (3) possible explanations for the differences in scores on state achievement tests and National Assessment of Educational Progress (NAEP) tests. When Congress gave the NAEP Governing Board authority for voluntary test development, it altered the relationship between the Governing Board and the Department of Education. Department of Education assistance was not used in developing the test contract, as had been typical of NAEP tests in the past. A review of the test development contract suggests that the cost to the federal government for developing one complete set of tests would be $15 million, and could go as high as $96 million for a complete set of tests for grades 4 and 8 if all students participate. A number of explanations have been proposed for the differences between state and NAEP results, but most center on differences in definitions of proficiency, differences in the kinds of tests administered, and differences in the kinds of students taking the tests. Appendixes present the study questions, a memorandum of understanding between the Department of Education and the Governing Board, the NAEP response to outside recommendations, and comments from both agencies. (Contains three tables and one figure.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

TM

ED 423 244

United States General Accounting Office

GAO

Report to the Honorable William F. Goodling, Chairman, Committee on Education and the Workforce, House of Representatives, and the Honorable John Ashcroft, U.S. Senate

June 1998

STUDENT TESTING

Issues Related to Voluntary National Mathematics and Reading Tests

TM028973

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

GAO/HEHS-98-163

2

GAO

United States
General Accounting Office
Washington, D.C. 20548

Health, Education, and
Human Services Division

B-278968

June 18, 1998

The Honorable William F. Goodling
Chairman, Committee on Education and the Workforce
House of Representatives

The Honorable John Ashcroft
United States Senate

In February 1997, the Clinton administration announced a plan to develop voluntary national tests for fourth grade reading and eighth grade mathematics. The tests were proposed as a way to provide parents and teachers with information about their students' performance relative to widely agreed-upon standards of what students should know and be able to do. Concerns have been raised, however, about the need for such tests and the potential for inappropriate federal influence on school curriculum, which is established by the states and localities.

Originally, the Department of Education had complete responsibility for the testing initiative. In November 1997, however, the Department's Appropriations Act transferred responsibility for developing the tests from Education to the National Assessment Governing Board (referred to as the "Governing Board"),[1] a board originally created as part of the Education Department in 1988 to set policy for a program called the National Assessment of Educational Progress (NAEP). In addition, the Congress directed the Department, including the Governing Board, to use no fiscal year 1998 funds to, among other things, pilot test, field test, or administer these tests.

The proposed voluntary national tests are designed to be based on NAEP. As part of your ongoing review of the proposed testing program, you asked us to provide you with information about three main issues:

- the relationship between the Governing Board and the Department of Education,
- costs of developing the national voluntary tests and procedures for hiring contractors, and
- possible explanations for the differences in scores on state achievement tests and NAEP tests.

[1]Departments of Labor, Health, Human Services, and Education and Related Agencies Appropriations Act, 1998. P.L. 105-78, sec. 307 (a), 111 Stat. 1467, 1505 (1997).

To address the first two issues and other specific questions (see app. I), we interviewed officials at both the Governing Board and the Department of Education, reviewed contract files, and analyzed contracts and other relevant documents. To determine reasons for differences between state and NAEP test results, we interviewed testing experts with the Governing Board, the Department, and the National Center for Research on Evaluation Standards on Student Testing and reviewed selected literature on testing standards. We also observed a Governing Board meeting and explored whether outside advisers to the voluntary national test specification process had written textbooks or educational materials for elementary and secondary education classrooms. We conducted our work between January and May 1998 in accordance with generally accepted government auditing standards.

Results in Brief

Since 1992, the Governing Board and the Department of Education have operated under a Memorandum of Understanding based on NAEP's authorizing legislation and designed to ensure the Governing Board's independence from its parent agency in fulfilling its responsibilities for NAEP, a national assessment program. In accordance with this memorandum, the Governing Board establishes program policies and standards, oversees contracts to develop test specifications, makes its own personnel decisions, obligates funds, and awards contracts. Because they share responsibilities for the program, however, the Department and the Governing Board collaborate extensively. Although the Governing Board helps in developing the Department's budget request for NAEP-related work, the Department has final authority over the request forwarded to the Office of Management and Budget (OMB).

When the Congress gave the Governing Board exclusive authority for the voluntary national test development contract in November 1997, it also altered the relationship between the Governing Board and the Department regarding this contract. Because it now had exclusive authority for this development contract, the Governing Board used no Department assistance in reviewing the contract as it had typically done for NAEP-related contracts. As a result of its review, the Governing Board substantially changed the test development contract. For example, under the Department's contract, the planned administration date of the first test was March 1999; under the Governing Board's contract, the planned administration date is March 2001. Another modification specified that the Governing Board, not the Department, would make all policy decisions as

specified in the original contract. Nor did the Governing Board help in developing the Department's budget request for voluntary national test development funds.

On reviewing the Governing Board's test development contract, we estimated that the cost to the federal government for developing one complete set of tests would be $15 million. The cost of implementing the fourth grade reading and eighth grade mathematics testing program has been estimated at up to $96 million if all fourth and eighth grade students in public and private schools participated. Who would pay for implementing the test has not been determined. For fiscal years 1997 and 1998, contracts awarded for voluntary national tests totaled about $17 million. Most of this amount—about $13 million—was for the 1997 contract for developing these tests; however, only about $1.7 million of this contract award had been paid to the contractor as of April 1998. In keeping with the congressional mandate to restrict the use of fiscal year 1998 funds, the only contract awarded with these funds was for a new congressionally mandated study by the National Academy of Sciences (NAS), which included examining the feasibility of comparing commercial and state tests with NAEP tests. In addition to those expenditures for which cost figures were available, the Department also spent an undetermined amount on distributing information about the national tests in fiscal years 1997 and 1998 by using web sites, press releases, articles in trade magazines, Department newsletters, and public presentations. Department officials said these dissemination activities are not directly related to developing the test and were part of the Department's ongoing function to educate the public about its program initiatives.

We heard various explanations for the differences in scores on the NAEP and state achievement tests; far fewer students achieve proficiency on recent NAEP tests than on state tests. Most explanations of the differences fell into three main categories: (1) differences in how the tests define proficiency and determine the threshold test score for proficiency (for example, one test may define proficiency as getting 66 percent of all items correct; another might define proficiency as scoring a minimum of 66 percent on each test section); (2) differences in the kinds of tests administered (for example, multiple choice versus problem solving); and (3) differences in students taking the test and when they take it. NAEP, for example, tests students with disabilities or limited-English proficiency; some state tests may not.

Background

Created by the Congress in 1988, the Governing Board provides independent, bipartisan policy guidance for NAEP, a congressionally authorized program in place since 1969. The Secretary of Education appoints the 25 voting members of the Governing Board to represent a wide range of legislatively prescribed categories, including two state governors from different political parties, a school board member, a school district superintendent, classroom teachers, a business or industry representative, and curriculum specialists. The Department's Assistant Secretary for Educational Research and Improvement serves as a nonvoting, ex officio member.

The Governing Board has joint responsibilities with the Department of Education for NAEP. The Governing Board sets policy for NAEP, and the Department's National Center for Education Statistics (NCES) oversees the development and implementation of the assessment according to the Governing Board's guidance. The NCES commissioner is responsible for NAEP operations and technical quality control. In November 1997, the Congress gave the Governing Board exclusive authority—separate from the Department—for the voluntary national test development contract.

The Governing Board and the Department Share Responsibility for NAEP

NAEP primarily tracks the educational achievements of students in grades 4, 8, and 12 in selected subject areas such as reading, writing, mathematics, and U.S. history.[2] The Governing Board and the Department share responsibility for developing and implementing each test.

For example, the Governing Board selects the subjects for each assessment cycle[3] and oversees the creation of the conceptual frameworks for the tests and the specifications for developing the tests. The frameworks identify subject-specific content areas and specify such test attributes as the cognitive skills to be tested and the type and number of test items in a particular content area. The frameworks are developed through a national consensus process that involves educators, academicians, policymakers, the general public, and others. NCES in turn hires contractors who use these frameworks to develop the actual assessments (which includes writing and pilot testing the test items and field testing the test booklet) and administer them. The Governing Board

[2]Thus far, the Governing Board has overseen the creation of assessment frameworks for reading, writing, mathematics, civics, arts, science, geography, and U.S. history. More frameworks for additional subject areas are planned.

[3]Since 1990, various state and national NAEP assessments have been administered every 2 years, according to a schedule determined by the Governing Board.

has final approval over each NAEP test item. Participation of students, schools, districts, and states in the assessment is voluntary.[4]

NAEP Gives State and National—Not Individual—Results

NAEP does not report on the performance of individual students or individual schools. It tests samples of students and reports the overall performance of student populations at the state and national levels. As required in its authorizing legislation, the Governing Board has developed student performance levels for each age and grade in each subject area tested. Starting with the 1990 mathematics assessment, the Governing Board has reported results using performance standards that define what students in a particular grade level should know and be able to do in each subject area tested. Student performance on the assessment is categorized as basic, proficient, or advanced.[5]

NAEP Standards Are High and Under Scrutiny

Experts generally consider NAEP assessments to be challenging, with high standards for achieving proficiency. In the 1994 NAEP reading test—the most recent one to measure reading achievement—only 36 percent of the nation's 12th graders were considered to be at or above the proficient level. Although the comparability of NAEP assessments with state assessments is being studied, state achievement tests usually result in a much larger percentage of students achieving proficiency.

Given the national significance of NAEP tests, the Governing Board's process for developing the performance standards has been widely scrutinized and evaluated. Reviews of the process have been done by NCES, research firms under contract either to the Governing Board or the Department, and other independent groups. For example, we previously studied the Governing Board's 1990 standard-setting process,[6] and the

[4]In 1994, 39 states participated in the fourth grade reading state-level assessment and met the minimum school participation guidelines for public schools. In 1996, 40 states participated in the eighth grade state-level mathematics assessment and met the minimum guidelines.

[5]"Basic" denotes partial mastery of prerequisite knowledge and skills for proficient work, "proficient" denotes solid academic performance that demonstrates competency in challenging subject matter, and "advanced" denotes superior performance beyond proficient grade-level mastery.

[6]Educational Achievement Standards: NAGB's Approach Yields Misleading Interpretations (GAO/PEMD-93-12, June 23, 1993).

National Academy of Education (NAE) studied mainly the 1992 process.[7] Both of these reviews disagreed with several aspects of the Governing Board's approach, which used a well-known standard-setting process with some novel modifications. The Governing Board considered these views along with several other diverse points of view[8] and changed its standard-setting process but kept its basic approach. The Governing Board plans to use the same approach for setting performance standards for the voluntary national tests as for the NAEP assessments.[9] (App. III describes the Governing Board's response to NAE's and our recommendations in more detail.)

Differences Between NAEP and Proposed Tests

The plan announced by President Clinton to develop voluntary national tests for fourth grade reading and eighth grade mathematics would be based on NAEP test frameworks. This means, among other things, that broad consensus would be obtained on the tests' subject matter and approach. The proposed tests would also differ from NAEP in several key respects, however, including providing individual results (see table 1 for a comparison of differences between NAEP and the proposed national tests). In addition, like NAEP, the proposed tests would have voluntary participation. The administration has announced that it would administer the first tests in March 1999.

[7]Lorrie Shepard, Robert Glaser, Robert Linn, and George Bohrnstedt, Setting Performance Standards for Student Achievement, A Report of the National Academy of Education Panel on the Evaluation of the NAEP Trial State Assessment: An Evaluation of the 1992 Achievement Levels, National Academy of Education (Stanford, Calif.: Stanford University, 1993). NAE is composed of 125 scholars and education leaders who promote scholarly inquiry and discussion about the ends and means of education in the United States and abroad. NAE has been in existence for about 33 years.

[8]See Proceedings of the Joint Conference on Standard Setting for Large-Scale Assessments of the National Assessment Governing Board (NAGB) and the National Center for Education Statistics (NCES), Volume II, and Executive Summary, Volume I, prepared by Aspen Systems under contract to the Governing Board and NCES (Washington, D.C.: Oct. 1995).

[9]The Governing Board further clarified how this will be done: "[T]he contract and specifications approved by the Governing Board require that the proposed tests be designed to show whether individual students can meet NAEP standards for basic, proficient, or advanced performance. This will be done to link various forms of the tests. No new standard-setting is anticipated. However, the Board may make some modifications or amplifications in the written descriptions of achievement levels to improve the clarity and usefulness of the performance standard." See app. IV.

Table 1: Key Differences Between NAEP and the Voluntary National Testing Proposal

Characteristic	NAEP	Voluntary national testing proposal[a]
What grades are tested	Primarily students in grades 4, 8, and 12	Students in grades 4 and 8
What subject areas are tested	Selected areas determined by the Governing Board	Reading (grade 4) and mathematics (grade 8)
Which students are tested	Random sample of students	All students in participating communities; but participation of individual students by law is voluntary
Extent of testing	Each student is tested on only some test items, with 45 to 60 minutes of testing per student	Each student is tested on all items, with up to 90 minutes of testing per student
Nature of participation	Voluntary on the part of every student, school, district, and state	Voluntary on the part of every student, school, district, and state
How results are reported	Individual scores not available; results are aggregated into a state or national score	Individual scores available and results can be aggregated for schools, districts, and states
Availability of test	Some test items released to the public, but most are not (as a result, parts of the test are used in subsequent years)	Test released to the public (as a result, a new test must be used for each test cycle)

[a]In general, these elements were initially proposed by the Department of Education but have not all been completely approved by the Governing Board since it acquired the test development contract, a Board official said. Although the Board has approved the test grades and subject matter and the extent of testing, it has made no decision about the type of communities participating in the test, the extent to which tests will be released to the public, or the level to which test scores will be aggregated. Such decisions are pending the Board's review and any further congressional action, according to the Board official.

Concerns About Proposed Tests

The voluntary national test proposal has raised many concerns. Among the concerns are that the tests (1) were not needed given the number of tests already being administered, (2) were being developed too quickly without allowing for enough congressional debate and consideration, (3) might result in inappropriate federal influence on state and local curricula, (4) would discriminate against students with disabilities and limited-English proficiency, and (5) might be used inappropriately to base federal aid received by a state on testing performance.

Partially in response to these concerns, the Congress took several steps to change the testing proposal and the timetable. It passed the Department of

Education's Appropriations Act for fiscal year 1998, which transferred authority for the contract to develop the national tests from the Education Department to the Governing Board (P.L. 105-78, sec. 307 (a)). The Department had already awarded the test development contract to the American Institutes for Research (AIR) in August 1997.[10] The Appropriations Act required the Governing Board to review this contract and change it or negotiate a new one as necessary. The act also prohibits the Department from spending fiscal year 1998 funds to field test, pilot test, administer, distribute, or implement any national tests (sec. 305 (a)). The act does not specifically prohibit using fiscal year 1998 funds, however, to develop test specifications or write test items. The act does prohibit the federal government from requiring states or districts to participate in any aspect of the test; it also prohibits the federal government from requiring any student to take the test (sec. 310 (a)). Finally, the act mandates NAS to conduct some studies on the comparability of NAEP assessments with state tests and commercial tests (sec. 306 (a)).

Governing Board's Relationship With the Department Depends on Whether the Focus Is on NAEP or Voluntary National Tests

The relationship between the Governing Board and the Department depends on whether the focus is on NAEP or the proposed voluntary national tests. To clarify this relationship regarding NAEP matters, the Department and the Governing Board signed a 1992 Memorandum of Understanding that while providing for collaboration, seeks to ensure the Governing Board's independence from the Department for its NAEP functions, powers, and duties. The memorandum, for example, calls for the Secretary to delegate to the Governing Board the authorities to hire personnel and enter into contracts. Funds for the Governing Board's operating budget and NAEP-related contracts come from the Department's Office of Educational Research and Improvement's (OERI) appropriations for NAEP tests. The Governing Board helps in developing the Department's budget request for these funds.

Passage of the Department's Appropriations Act in November 1997 changed the relationship between the Governing Board and the Department by transferring exclusive authority for the voluntary national test development contract from the Department to the Governing Board. Since that time, both Governing Board and Department officials have indicated that congruent with the act, the Department has had almost no involvement with the Governing Board's handling of this contract. Thus,

[10]AIR is a 50-year-old not-for-profit organization that specializes in conducting educational research, including large-scale assessments, policy analysis, and program evaluation.

unlike the two bodies' typical collaboration on NAEP matters, almost no collaboration takes place on voluntary national test matters, officials report. The Department funds the Governing Board's voluntary national test development contract costs with moneys drawn from another OERI program, the Fund for the Improvement of Education (FIE). The Governing Board has not helped in developing the Department's budget request for FIE funds.

Memorandum Determines Relationship Regarding NAEP Matters

The 1992 Memorandum of Understanding is based on the 1988 legislation authorizing NAEP.[11] The memorandum calls for the Department to delegate to the Governing Board the authority to, among other activities, appoint excepted service personnel; obligate funds; obtain the services of experts and consultants; and make, enter into, and perform contracts. The memorandum also calls for the Department to provide funding and administrative support services to the Governing Board.

Personnel Matters

In addition to its 25 members, the Governing Board has a staff of 10: 6 professional staff with technical expertise, assisted by 2 support and 2 clerical staff. Recently, the Governing Board hired three part-time, temporary consultants to help with its new responsibilities for the voluntary national tests, a Governing Board official reported.

In accordance with its authority, the Governing Board makes all staff hiring decisions, according to Board officials. In addition, Board officials said they determine positions' specifications, classifications, and salaries. For administrative purposes, however, Governing Board staff are considered Department of Education employees, and, in some cases, the Governing Board uses the Department's administrative services to advertise and conduct initial screenings for these staff positions, according to Board officials. Both Governing Board and Department officials told us that no Department employee or appointee has ever been loaned or detailed to work for the Governing Board. Department staff are available, however, when the Governing Board requests assistance.

Regarding the Board, the Secretary of Education continues—although it is not required—to select the Board members from a list of candidates nominated by the Board, rather than select members without the Board's help. The original 1988 NAEP legislation required the Secretary to choose from the Governing Board's nominations; the 1994 NAEP legislation

[11]A copy of the Memorandum of Understanding appears as app. II.

removed this requirement, although the Secretary opts to use the Board's nominations.

Contracting and Related Matters

The Governing Board has full authority to hire contractors for NAEP-related work. The Department does not approve these contracts in any way, according to Governing Board and Department officials. The Governing Board has three types of contracts for NAEP work—two are for technical contracts that develop test frameworks or set performance standards, and the third is for operational contracts that arrange the logistics of meetings. In developing these contracts, the Governing Board involves some Department personnel, but these staff have no approval authority for the contracts, Governing Board officials said. For example, the Board has asked NCES to participate with outside experts in a peer review of the technical contracts.

Coordination between the Board and the Department is important in developing NAEP contracts, according to officials of both bodies. For example, to ensure appropriate time lines for the Governing Board contractor to develop achievement levels for the civics test administered in 1998, the Board had to determine when the Department's contract work for developing the civics test items would be completed. In addition, department officials said that in the past, the Department has provided the Board with training and assistance on using computer software related to contract documentation. The Governing Board has not requested such assistance, however, in the past 2 years.

Collaboration on NAEP Matters

The Memorandum of Understanding has provisions for collaboration on NAEP matters between the Governing Board and the Department. Officials of both bodies said that collaboration on NAEP matters was important and perhaps essential for the Department to implement NAEP according to the Board's policy and procedures. This collaboration does not affect the Board's independence from the Department, according to Board officials. The Governing Board and Department officials described the following examples of their collaboration on NAEP matters:

- By participating as a nonvoting Board member, the Department's ex officio member helps to ensure that the Department understands the Governing Board's policy guidance on NAEP frameworks and achievement levels. Similarly, a Governing Board staff member attends NCES meetings on test development to convey the Board's policies.

- Regarding formulating NAEP policy for test frameworks or achievement levels, the Governing Board seeks technical advice from a broad range of experts, some of whom include experts with NCES and its contractors.
- The memorandum calls for the Department and the Board to seek mutually satisfactory resolutions when NCES disagrees with the Governing Board's policy guidance. In a few cases, the Secretary of Education has been the final arbiter of conflicts, according to Governing Board and Department officials, according to Department officials.

Funding for Operating Budget and NAEP Contracts

The funding for the Governing Board's operating budget and NAEP-related contracts comes from OERI's appropriations for NAEP assessments. From these appropriations, the Governing Board was allocated about $2.9 million and $3.5 million in fiscal years 1997 and 1998, respectively. The Board helped develop the Department's budget request for the Board's part of NAEP's appropriations.

Using Department guidance, the Governing Board develops its annual budget request on the basis of the estimated cost of its staff salaries, Board member expenses, and NAEP contracts and submits it to the Department, Governing Board officials said. Department budget officials said this guidance consists of the typical instructions required by OMB or congressional directives. The Department reviews the Governing Board request, makes adjustments, and includes the request in its total agency request to OMB. The amount of the Department's request may differ from the amount the Governing Board submitted, and the Board is generally not involved in the Department's negotiations with OMB during OMB's review. For example, the Governing Board submitted a request of $4.8 million for fiscal year 1999, and the Department reduced this request to $3.4 million, according to officials of both bodies. After reevaluating the Governing Board's funding needs, however, the Department later petitioned OMB to increase the Board's budget to $4 million, which OMB did.

Except for Funding, the Department Has Had Little Involvement in the Governing Board's Voluntary National Test Work

Governing Board and Department officials view their relationship regarding the voluntary national test work differently from their collaboration on NAEP, reflecting the Appropriations Act provisions. After the Board received exclusive authority for the test development contract, officials of both bodies said they discussed the mechanics of smoothing the transfer of authority, the legal availability of funds to pay contractors for pilot and field tests, and a legal matter involving students with disabilities and limited-English proficiency. Other than these instances, officials have not collaborated on voluntary national test work, they said.

Department Not Involved in Contract Review

Upon assuming exclusive authority for the voluntary national test development contract, the Governing Board was required to review the contract the Department had already awarded and modify it or negotiate a new contract as necessary. To review the contract, the Governing Board formed a special committee of its Board members that did not include any Department officials. On the basis of the special committee's recommendations, the Governing Board approved some major changes to the contract, including the timing of pilot testing, field testing, and test administration. For example, under the Department's contract, the first test was to be developed in time to be administered in March 1999; under the Governing Board's contract, the test is to be developed in time to be administered in March 2001. In addition, the Board's contract changed responsibility for contract development policy decisions from the Department to the Governing Board, in keeping with the Appropriations Act.

The act also directed the Governing Board to determine, among other things, whether the test development process and test items consider the needs of disadvantaged, limited-English proficient, and disabled students. Because the Board has no legal counsel of its own, the Board sought advice from the Department and other sources about the access of such students to voluntary national tests, according to a Board official.

Funding for the Governing Board's Voluntary National Test Development Contract

The Department funds the Governing Board's voluntary national test development contract costs with moneys from FIE, a discretionary account that gives the Department the authority, under 20 U.S.C. 8001(a), to support nationally significant programs and projects to improve education. As of April 1998, the Department has paid the contractor a total of about $1.7 million, all with fiscal year 1997 FIE funds. The Department has allotted about a third of these payments ($540,000) to the Governing Board for payment to the contractor. The Governing Board has not been involved in developing the Department's budget request for FIE funds, Department and Governing Board officials said.

Fiscal Years 1997 and 1998 Contract Awards and Expenditures and Future Cost Estimates

Except for the test development contract specifically transferred to the Governing Board, the Department still has responsibility for other contracts related to the proposed voluntary national tests. These other contracts cover such activities as developing test specifications based on NAEP frameworks,[12] comparing NAEP with other state and commercial tests,

[12]The Department's test specifications contract was completed before Nov. 1997, when the Board received authority for the test development contract.

and evaluating the voluntary national test items. The Appropriations Act requires these last two activities.[13] The Department awarded voluntary national testing contracts totaling about $17 million for fiscal years 1997 and 1998. Most of this amount obligated the test development contract that was subsequently transferred to the Governing Board. Contract awards represent obligated funds; payments of such funds to contractors are termed "outlays," a Department budget official said. Outlays and other expenditures for voluntary national tests in fiscal years 1997 and 1998 totaled almost $3 million as of April 1998. (See table 1 for a summary of contract awards and expenditures.) Under the Governing Board's revised contract terms, we estimated the cost for developing one complete set of voluntary national tests is $15 million, which the federal government is expected to pay. The Department determined that costs to implement the fourth grade reading and the eighth grade mathematics testing program in a year could total up to $96 million, depending on the number of students tested. Who would pay these costs has not been determined.

[13]The Department signed the test evaluation contract before the passage of the Appropriations Act; however, Department officials said that the contract fulfills the act's mandate (sec. 308) for NAS to evaluate the test items developed for the voluntary national tests.

Table 2: Education Department Contract Awards and Expenditures for Voluntary National Tests, Fiscal Years 1997 and 1998

Purpose and type of expenditure[a]	Contract award amount or expense[b]	Contract outlay and other expenditures[c]	Contractor or recipient	Funding source
Fiscal year 1997				
Task order contract for test specifications (awarded 5/1/97)	$960,473	$882,655	MPR Associates	FIE
Contract for test development (awarded 8/18/97)	13,035,848	1,718,728	AIR	FIE
Contract for test evaluation (awarded 9/30/97)	1,238,800	260,988	NAS	FIE
Task order contract for meetings with experts and transcripts (awarded 2/14/97)	26,439	17,858	Professional and Scientific Associates	FIE
Expenses related to August 1997 conference with state education officials	28,720	28,720	Loews L'Enfant Plaza Hotel, American Express	OERI salaries and expenses funds
Fiscal year 1997 totals	15,290,280	2,908,949	Not applicable	Not applicable
Fiscal year 1998				
Contract for test equivalency scale, forms, and use studies (awarded 2/3/98)	1,800,000	80,270	NAS	FIE

[a]Information in this table is based on contract records and expenditure summaries provided by the Department of Education and the National Assessment Governing Board.

[b]With the exception of the test development contract, these contracts and expenses are under the jurisdiction of the Department. The test development contract, originally signed by the Department in Aug. 1997, was transferred to the Governing Board in Nov. 1997.

[c]These outlays were made as of Apr. 1998. The only outlays made by the Governing Board account for about $540,000 of the $1.7 million outlaid for the test development contract. The Department made all other outlays.

Contract Awards in Fiscal Years 1997 and 1998

In fiscal year 1997, the Department awarded contracts worth about $15 million for voluntary national test activities. Preparing for these contracts, which began just before the President's announcement of the voluntary national test initiative, represents the Department's first

spending for the initiative, Department officials said.[14] The Department based its authority for funding the activities of the national testing program on the general authority under FIE.[15] The total $15 million awarded to the various contracts came from FIE funds, which are administered by OERI.[16]

Most ($13 million) of the fiscal year 1997 award amount was for the test development contract awarded to AIR. As of April 1998, only about $1.7 million had been spent on this contract, however, because of a partial stop-work order first initiated by the Department and later extended by the Governing Board.[17] The Department has carried forward the obligated balance, about $11.3 million, of the fiscal year 1997 FIE funds to pay the contractor's costs, now that the stop-work order has been lifted and the contractor has resumed working, a Department budget officer reported.

Of the three other fiscal year 1997 contracts, whose award amounts totaled about $2 million, the contractors have received only about $1.2 million. In addition, the Department used $28,720 of OERI salary and expense fund moneys to pay for invitational travel, meeting space, use of video and audio equipment, and a working lunch for nonfederal participants during the Secretary's conference on the voluntary national tests with state education officials.

For fiscal year 1998, the Department had planned to make another $15 million in FIE funds available for voluntary national test activities. As mandated by the Department's 1998 Appropriations Act, the Department awarded $1.8 million of these funds to NAS to conduct studies on test equivalency scales and other issues. For this contract, NAS will study (1) the feasibility of developing equivalency scales for comparing test scores on different commercial tests and state tests with each other and with NAEP scores and (2) the steps needed to ensure that test results are

[14]Because the voluntary national tests use NAEP as a foundation, funds spent on NAEP (for example, about $33 million appropriated in fiscal year 1997 and about $35 million in fiscal year 1998) relate somewhat to voluntary national tests, although they are technically not part of the development costs.

[15]Under this authority, the Department is further authorized to (1) support the development of model strategies for assessing student learning, (2) develop and evaluate strategies for integrating instruction and assessment so that teachers and administrators can focus on what students should know and be able to do at particular grade levels, (3) support nationally significant programs and projects to improve the quality of education, and (4) support activities to raise standards and expectations for all students' academic achievement.

[16]FIE's budget request for fiscal year 1997 was $40 million; the overall OERI appropriation, including FIE, was $293.1 million.

[17]The Department issued a partial stop-work order in Sept. 1997 to save the government from spending funds for work that may have to be redone if the Congress transferred the contract to the Governing Board. The Governing Board extended the stop-work order until Feb. 1998 to allow it time to review the contract.

BEST COPY AVAILABLE

not used in a discriminatory way for student promotion, tracking, or graduation and that test forms will accurately gauge student achievement in reading and mathematics. As of April 1998, the Department had spent about $80,270 on this NAS contract.

None of the remaining fiscal year 1998 FIE funds available for voluntary national test activities has been obligated, a Department budget officer reported. As identified in its December 1997 spending plan, the Department had planned to use $13 million from the fiscal year 1998 FIE funds for the second year of the test development contract.[18] Contract delays, a congressional prohibition on using fiscal year 1998 funds for some test development activities, and contract modifications, however, will probably reduce the amount of fiscal year 1998 FIE funds needed for the second year of the contract, a Department budget officer said. The Department is waiting for the Governing Board to communicate the amount needed for the contract's second year and will allot the amount requested, the budget officer explained. The budget officer also said that no other FIE or Department funds appropriated in fiscal year 1998 are being spent on development activities for the proposed tests.

In addition to those expenditures for which a cost figure was available, the Department spent an undetermined amount when distributing information about the national tests in fiscal years 1997 and 1998. The Department has used various means to inform the public about the tests, including web sites, press releases, articles in trade magazines, Department newsletters, and presentations. Department officials said these activities are not directly related to developing the test but were part of the Department's ongoing function to educate the public about Department initiatives. Department officials had no estimate of the costs of these activities, they said, because the costs were relatively small and an integral part of the larger costs of salaries, supplies, and services.

Test Cost Estimates for Future Years

The Department originally estimated that the total cost for the original 5-year contract would be about $64 million. During the 5-year term of the contract, the contractor would have developed four complete test sets and

[18]The Department's appropriations for fiscal year 1998 specifically excluded the use of fiscal year 1998 funds for some but not all of the activities covered under the 1997 test development contract. For example, the contractor could use fiscal year 1998 funds to write test items, but it could not conduct field or pilot testing. In any event, no fiscal year 1998 funds have been spent on the contract for developing voluntary national tests, a Department budget official reported.

part of a fifth.[19] On the basis of these figures, the average cost of developing one complete test set is about $14 million.[20] The Governing Board's revised test development contract also covers a 5-year period. To ensure that pilot testing and then field testing take place in the same month of subsequent school years, however, the Governing Board increased the time for developing one complete test set from about 15 months to about 34 months.[21] During the 5-year period, the contractor would be required to develop three test sets for a total award amount of about $46 million. As a result, the estimated average cost of developing a complete set of tests under the revised contract is about $15 million, which the federal government is expected to pay.

To help estimate the implementation costs of administering the test to students, a Department official asked some states what they spent administering state or commercial assessments. Per pupil costs ranged from $10 to $12. Because the methods used to administer and score voluntary national tests may differ from the various state methods, however, these figures may not be appropriate for comparison purposes, according to Department officials. Nonetheless, assuming it cost $12 to test each of the estimated 8 million students eligible to take the test, the Department estimated that implementing the fourth grade reading and eighth grade mathematics testing program would cost up to $96 million a year.[22] The administration had originally proposed that the federal government would reimburse states and localities for costs in the first year of full testing. The Governing Board now has to address this issue, however, Department officials said. Board officials believe the Board is legally responsible only for developing the assessment—not implementing it. Who would pay for implementing these tests is pending the Congress' decision, according to the Board.

[19]In the 5-year period covered by the original contract, the contractor would have been required to complete all activities needed to develop and finalize four sets of voluntary national tests for mathematics and reading, one set for each year the test was to be administered in years 1999 to 2002. Each set of tests required about 15 months to complete and included writing new test items along with the pilot and field testing of these items. To meet the planned schedule for administering the tests, the contractor would have to develop some test sets concurrently.

[20]A test set includes several different but equivalent tests for each subject. The set includes, for example, a copy used for the actual test, a backup copy in case security has been breached, a copy reserved for research purposes, and a sample copy released to the public.

[21]The pilot and field testing of students is scheduled to take place a year apart in the same month of the school year to ensure that tested students have about the same amount of prior knowledge. The contractor is expected to use the results of pilot testing to modify the test items as needed before conducting field testing.

[22]This estimate is based on the assumptions that (1) all fourth and eighth graders in public and private schools will take the voluntary national tests in 1999 and (2) such students will number 4 million in each of these grades in 1999. In 1995, the most recent year for public school enrollment figures, about 3.4 million students were enrolled in each of grades 4 and 8.

Contractor Selection

The contractor selection process for the 3 fiscal year 1997 contracts directly related to the national tests involved developing test specifications,[23] developing the test, and evaluating the voluntary test initiative.[24] The Department used a competitive procurement process for all three contracts but was not able to obtain more than one proposal for each. The Department mainly used its own staff to evaluate the proposals' technical and cost aspects. For the test development contract, the Department also formed a technical review panel of federal and nonfederal experts.

In addition, none of the individuals selected as advisers for developing test specifications, including members of the National Test Panel and the three advisory committees,[25] had produced textbooks or materials used in primary and secondary school classrooms, according to our review.

Test Specifications Contract

In 1996, the Department established two competitive task order contracts with two firms, AIR and MPR Associates. Under the terms of these contracts, each firm agreed to compete for task orders that would provide technical and analytical support to OERI on an ad hoc basis. In 1997, the Department determined that the voluntary national test specifications work could be done as a competitive task order under the competitive task order contracts with AIR or MPR Associates. The Department had determined, however, that firms could not compete for both the test specifications task order and the test development contract because competing for the test specifications task order would reveal information that would give a firm an unfair advantage in competing for the test development contract—in violation of the Federal Acquisition Regulation

[23]The Governing Board later replaced the test specifications developed for the voluntary national tests for this contract with a set of specifications that are closely based on the assessment framework the Board previously adopted. (See app. IV.)

[24]We did not review the selection process for the fiscal year 1998 contract because the Congress had specifically mandated the awardee, and we did not review the task order contract that established administrative arrangements for meetings with experts because it was small and not directly related to the test.

[25]The Reading Committee, composed of teachers, scholars, teacher-educators, and reading specialists, developed a set of recommended item and test specifications for fourth grade reading. The Mathematics Committee, composed of mathematics teachers, supervisors, and educators and university mathematicians, developed recommended item and test specifications for eighth grade mathematics. The Technical Advisory group, composed of measurement and assessment experts, assisted the other panels to ensure that the item and test specifications would result in fair, valid, and reliable measures.

(FAR).[26] Because AIR intended to compete for the test development contract, it did not compete for the test specifications task order.

The Department, therefore, entered into sole-source negotiations with MPR as the only firm available to perform the test specifications task order under the competitive task order contract, said Department officials. The Department awarded MPR the task order under MPR's existing task order contract in May 1997.

To help reach a national consensus on the appropriateness of the specifications, the contract called for involving a wide range of experts. The contractor formed a National Test Panel to make the final recommended item and test specifications. This panel included individuals from key constituencies and policy-making groups for elementary and secondary education at the state and local levels. To facilitate the panel's work, the contractor also established three committees.

According to our review of databases for documents written by any of the 56 panel and committee members, some members had written technical papers, analyses of education issues, or resource books for teachers.[27] No members, however, had written textbooks or educational materials for elementary and secondary classroom instruction in grades kindergarten through 12.

Test Development Contract

The Department contracted the development of the voluntary national tests by competitive negotiated acquisitions, according to Department officials and a review of contract records. Department officials said, however, that the contracting process was more open and public than usual. For example, the Department held a series of public meetings for market research purposes, to inform the public, and to gather advice for the project. The Department also publicized a draft statement of work and the request for proposal (RFP) through both the Commerce Business Daily and the Department's web site in spring 1997. The Department expected

[26]The FAR requires contracting officers to avoid, neutralize, or mitigate conflicts of interest before awarding a contract. The Department determined that an organizational conflict of interest would result if it permitted the same firm to compete for both the test specifications and test development contracts (FAR §§9.504(2), 9.501).

[27]The four databases we searched were the Educational Resources Information Center of the U.S. Department of Education; the Library of Congress catalog; Books in Print, a comprehensive source of books published in the United States, including bibliographic records from over 44,000 publishers; and the IAC Business A.R.T.S., a scholarly business database designed to help on-line searchers identify current research, analysis, trends, and expert opinion in a variety of disciplines.

the increased publicity to produce a number of proposals, Department officials said.

Although about 90 offerors requested copies of the RFP and an undetermined number of individuals downloaded it from the web site, only one firm, AIR, submitted a proposal by the June 1997 closing date. AIR's proposal included seven subcontractors, who had extensive collective experience in developing, field testing, and publishing standardized tests.[28]

The Department formed a technical review panel to determine if AIR's proposal met the RFP's criteria and recommended whether it was acceptable to perform the RFP's tasks. Panel members included seven Department staff representing a range of expertise in assessment, legal issues, reading, mathematics, and other areas and two outside experts, one representing a district-level perspective in math assessments and the other experienced in state-level assessments. A Department individual who had been involved in the test initiative served as the panel's technical adviser. The panel determined that the proposal could be made acceptable through negotiation. The Department's contract officer reviewed AIR's and the subcontractors' cost information. Before negotiating with AIR, the contract officer requested advice from Department staff who knew about the testing initiative's activities that occurred in public hearings and high-level policy meetings. Accordingly, the Department's Director of the Voluntary National Test Initiative and an OERI budget officer were briefed on maintaining procurement integrity and made procurement officials. These two individuals helped formulate the Department's negotiation strategy. After the Department negotiated with AIR to clarify technical, business, and cost issues, AIR provided its best and final offer, which the Department accepted in August 1997.

When the Governing Board received exclusive authority for the contract, the Board had 90 days to review the contract and accept, modify, or terminate it. The Governing Board ultimately decided to modify the contract and submitted a revised statement of work to AIR. The Governing Board and AIR signed a revised contract in February 1998.

Test Evaluation Contract

Although OERI originally requested that the test evaluation contract be awarded on a sole-source basis to NAS, the Department's contract officer recommended a search for other qualified sources. Accordingly, the

[28]The subcontractors were Riverside Publishing; Harcourt Brace Educational Measurement; Westat, Inc.; California Test Bureau; National Computer Systems, Inc.; Educational Testing Service; and Council for Basic Education.

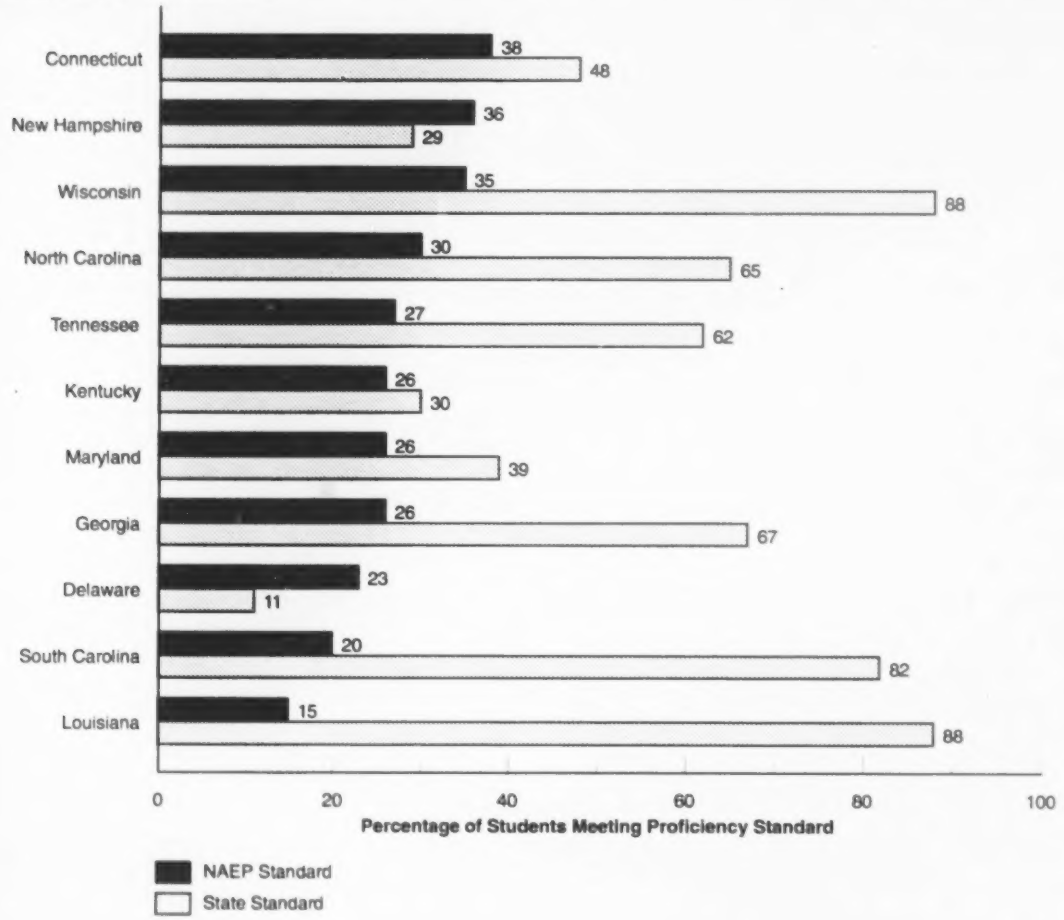
Department issued a sources-sought announcement for this work in the Commerce Business Daily in July 1997. No firms responded to the announcement, however. As a result, the Department began sole-source procurement with NAS. After the Department's contracting personnel conducted technical, business, and cost reviews of NAS' proposal, the Department eventually awarded this contract to NAS for about $1.2 million in September 1997.

Several Explanations Offered for Differences Between NAEP and State Test Results

Some controversy has developed over NAEP's framework for voluntary national testing because of the differences in results between NAEP and some state-level proficiency tests. Many states conduct their own tests to measure student proficiency using state-developed tests or those from a private testing service adapted for states' use. In his September 1997 testimony, the Secretary of Education presented data from a study that compared test results in 11 states (see fig. 3).[29] In most cases, the percentage of students meeting the NAEP proficiency standard is much lower than the percentage meeting the proficiency standard for the state examination. The study's author suggested that one important reason for this difference is because most state proficiency standards appear to be set at a lower level than NAEP's. Our discussions with several testing experts indicated that different proficiency standards could be one explanation, but other explanations are also possible.

[29]Mark D. Musick conducted this study in his capacity as president of the Southern Regional Education Board. Mr. Musick is also a Governing Board member.

Figure 1: Smaller Percentage of Students Meeting the Proficiency Standard on NAEP Fourth Grade Reading Test Compared With Most State Tests, 1994



Note: State results for New Hampshire, Wisconsin, Maryland, Georgia, Delaware, South Carolina, and Louisiana are based on reading tests of third rather than fourth graders.

Source: This figure is based on data from "Setting Education Standards High Enough," by Mark D. Musick, President, Southern Regional Education Board, July 1996. We confirmed that NAEP's scores reported in the figure were the same as those reported by NCES for the 1994 NAEP reading assessment.

Defining Proficiency Varies

In his testimony in September 1997, the Secretary of Education highlighted the differences in proficiency scores between NAEP and state tests used by the 11 states involved in the study. The study's author suggested that the differences may be partly attributed to NAEP's assessments having a higher proficiency standard. That is, a student might have the same correct response rate on two comparable tests, but the expectations of what amounts to a proficient score would be higher for one test than for the other.

Experts we spoke with agree that differences in tests' proficiency definitions can produce differences in the number of students judged to be proficient. For example, one test expert described to us how two states using about the same test had different expectations of the amount and type of content the student had to master. One state defined proficiency as correctly answering 66 percent of the items in each major content area of the test, while the other state required an overall correct response rate of 66 percent. Under the latter approach, a student might score below 66 percent in some content areas but still get a proficient score, while a student scoring the same in the former state would not pass the test.

Furthermore, experts explained that even if NAEP and states had common definitions of proficiency, different approaches for setting the proficiency threshold score at or above which a student is scored proficient could lead to different scores being designated as proficient. For example, one approach, called "item-centered," establishes a proficiency score, called a "cut" score, for each performance level based on an evaluation of the test items. NAEP uses an item-centered approach to determine its cut score. Another approach establishes cut scores on the basis of actual test performances by students judged by other means to be at a certain competency level.

Test Instruments Vary

Another possible reason for the difference in test results, experts we spoke with said, may be the actual tests. NAEP's assessments may differ from an individual state's test by the amount and type of content covered, the format (such as the extent to which a test uses multiple-choice or open-ended questions), and the type and extent of cognitive skills tested (such as the extent to which the student is asked to simply recall facts or apply logic in solving problems).

A test's relationship to a particular curriculum being taught could also affect scores, according to experts. One expert suggested that compared

with NAEP's assessment, some state tests may more closely relate to what was actually being taught. In contrast, NAEP assessments relate less to state school curricula, this expert said.

Student Profiles Vary

Although not generally considered an important reason for explaining the differences in NAEP and state test results, differences in students taking the test and when they take it may affect the results, according to experts. For example, experts we spoke with cited students who take the NAEP tests include those with disabilities and those with limited-English proficiency; state tests may not include such students. In addition, a student taking a test early in the school year may have less knowledge than one taking it later in the year.

Agency Comments

We provided both the Governing Board and the Department of Education the opportunity to comment on a draft of this report. The Governing Board said that the report was fair, balanced, and accurate. It also suggested several specific technical changes that we incorporated as appropriate. The full text of its comments appears in appendix IV. The Department of Education did not directly comment on the report but said that the Governing Board has made good progress and should continue to move forward in its efforts to develop national tests. The full text of the Department's comments appears in appendix V.

We are sending copies of this report to appropriate congressional committees and members of the Congress, the Secretary of Education, the National Assessment Governing Board, and other interested parties. Please call me at (202) 512-7014 or Eleanor L. Johnson, Assistant Director, at (202) 512-7209 if you or your staff have any questions.

Cornelia M. Blanchette

Cornelia M. Blanchette
Associate Director, Education and
Employment Issues

Contents

Letter		1
Appendix I Specific Study Questions		28
	Relationship Between the Governing Board and the Department of Education	28
	Costs of Developing the National Voluntary Tests and Procedures for Hiring Contractors	29
	Possible Explanations as to Why Scores on State Achievement Tests Differed From NAEP Test Results	29
Appendix II Memorandum of Understanding Between the Department of Education and the National Assessment Governing Board		30
Appendix III National Assessment Governing Board's Response to Outside Recommendations		33
	Background	33
	Overview of Our Recommendations and the Governing Board's Responses	34
	Overview of NAE's Recommendations and the Governing Board's Responses	34
	Perspectives on the Governing Board's Continued Use of an Item-Centered Approach to Standard Setting	35
Appendix IV Comments From the National Assessment Governing Board		39

Appendix V Comments From the Department of Education 41

GAO Contacts and Staff Acknowledgments 42

Tables

Table 1: Key Differences Between NAEP and the Voluntary National Testing Proposal 7
Table 2: Education Department Contract Awards and Expenditures for Voluntary National Tests, Fiscal Years 1997 and 1998 14
Table III.1: Our 1993 Recommendations and the National Assessment Governing Board's Actions Taken in Response 37

Figure

Figure 1: Smaller Percentage of Students Meeting the Proficiency Standard on NAEP Fourth Grade Reading Test Compared With Most State Tests, 1994 22

Abbreviations

AIR	American Institutes for Research
FAR	Federal Acquisition Regulation
FIE	Fund for the Improvement of Education
NAE	National Academy of Education
NAEP	National Assessment of Educational Progress
NAS	National Academy of Sciences
NCES	National Center for Education Statistics
NCEST	National Council on Education Standards and Testing
OERI	Office of Educational Research and Improvement
OMB	Office of Management and Budget
RFP	request for proposal

Specific Study Questions

Our study addressed three main issues: (1) the relationship between the Governing Board and the Department of Education, (2) costs of developing the national voluntary tests and procedures for hiring contractors, and (3) possible explanations for the differences in scores on state achievement tests and National Assessment of Educational Progress (NAEP) tests. Following is a list of specific study questions grouped by main issue.

Relationship Between the Governing Board and the Department of Education

1. What is the relationship between the Department of Education and the Governing Board? Specifically,

1.1. For each of the past 3 years, have any Department of Education employees or appointees been loaned, detailed, or otherwise made available to perform services for the Governing Board? If so, please list their names, titles, ranks, and months of service to the Governing Board.

1.2. What, if any, role does the Department of Education play in procurement for the Governing Board and NAEP?

1.3. Is the Department of Education involved in reviewing or approving contracts or any other activities or policies of the Governing Board? If yes, explain the specific way in which the Department is involved.

1.4. Is the Department of Education involved in the hiring of staff of the Governing Board and, if so, in what ways?

1.5. What role and influence, if any, does the Department of Education play or exercise in the distribution of annual appropriations to the Governing Board (for the activities of the Board and NAEP)? What role and influence, if any, does the Department of Education play or exercise in the establishment of the budget of the Board? Please provide any and all documentation of any agreements or memoranda of understanding?

2. How has the Governing Board responded to GAO reports and National Academy of Education analyses and recommendations?

Costs of Developing the National Voluntary Tests and Procedures for Hiring Contractors

3. How were contractors for national test development activities and individuals providing advice on the process selected?

4. Is the Department of Education using any fiscal year 1998 funds for activities related to developing the national tests?

5. How much money did the Department of Education spend on the concept or specific development of the national tests in fiscal year 1997 and fiscal year 1998 (1) before the testing initiative was announced and (2) after it was announced; and what are the details of these expenditures, such as the sources of funds, dates, and purposes of expenditures?

6. What is the Department of Education's estimate of the total cost of developing the proposed costs and the annual implementation costs that would be borne by (1) the federal government and (2) state and local governments?

7. What textbooks or educational materials have members of the national test panel, the technical advisory committee, or the math or reading committee authored? What, if any, federal funds were used to develop or produce these materials?

Possible Explanations as to Why Scores on State Achievement Tests Differed From NAEP Test Results

8. What are the alternative explanations for differences between student scores on state achievement tests and on the NAEP reading tests exhibited by Secretary Riley during his September testimony?

Memorandum of Understanding Between the Department of Education and the National Assessment Governing Board

UNITED STATES DEPARTMENT OF EDUCATION

THE SECRETARY

Memorandum of Understanding

The purpose of this Memorandum is to clarify the relationship between the U.S. Department of Education (the Department) and the National Assessment Governing Board (the Board). Particular, but not sole, attention is directed toward specifying how the Board may, to the maximum extent permitted by law, exercise its functions, powers and duties independent of the Secretary and the other offices and officers of the Department.

The Department and the Board recognize that authority for the Board derives from the authorizing legislation for the National Assessment of Educational Progress (NAEP). It is further recognized that the NAEP legislation contains requirements intended to guide the relationship between the Department and the Board. See 20 U.S.C. Section 1221e-1(i). Among these requirements, the NAEP legislation provides that:

- The Board shall formulate the policy guidelines for the National Assessment.
- The Secretary shall report to the Board at regular intervals on the Department's action to implement the decisions of the Board.
- Not to exceed 10 percent of the NAEP appropriation may be used for administrative expenses of the Board and to carry out its functions.
- In the exercise of its functions, powers and duties, the Board shall hire its own staff and shall be independent of the Secretary and the other offices and officers of the Department of Education.
- The Secretary may appoint, at the direction of the Board...not more than 6 technical employees to administer [20 U.S.C. Section 1221e-1(i)...under an excepted appointing authority].
- The Secretary and the Board shall ensure at all times that the membership of the Board reflects regional, racial, gender and cultural balance and that it exercises its independent judgment, free from inappropriate influences and special interests.
- As vacancies occur, new members of the Board shall be appointed by the Secretary from among at least three individuals who are nominated by the Board.

400 MARYLAND AVE., S.W. WASHINGTON, D.C. 20202-0100

- For the purposes of its administrative functions, the Board shall have the authorities authorized by the Federal Advisory Committee Act and shall be subject to the open meeting provisions of that law.

In furtherance of these requirements, the Board and the Department mutually agree to be guided by the following principles:

- That the Board will exercise its functions, powers and duties independent of the Secretary and the other offices and officers of the Department of Education to the extent permitted by applicable law.
- That the Department will make every reasonable attempt to implement the policy-setting actions of the Board.
- That in any instance where Board policy-setting actions are in conflict with other policies of the Department or the Administration, or when they cannot be accomplished, the Department and the Board will seek mutually satisfactory resolutions.
- That the Board and the Department will act to keep free and open communications between them, assuring that responsible officers are fully informed on a timely basis.
- That insofar as possible, within constraints of applicable Executive Branch procedures, the Department will facilitate communications between the Board and the Congress, the NAEP contractor, and the Executive Branch offices.

In furtherance of these principles, the Department and the Board agree to the following actions:

(1) Each fiscal year, the Secretary shall allot to the Board the Board's annual allocation, which will be available to the Board to perform its functions and duties, as determined by the Board. Such allotment shall generally be made on a quarterly basis. The Board shall comply with Departmental funds and personnel ceiling control procedures and requirements and provide such reports as the Secretary may reasonably require to account for the use and expenditure of such funds.

(2) The Secretary of Education shall, to the extent consistent with the Constitution, and applicable law and Executive Branch procedures, delegate to the National Assessment Governing Board, with provision for redelegation, the following authorities:

(a) the authority to appoint excepted service personnel as

BEST COPY AVAILABLE

provided under Section 406(i)(5)(C)(iii) of the General Education Provisions Act;

(b) the authority to obligate funds;

(c) the authority to sign notices for publication in the Federal Register and the Commerce Business Daily;

(d) the authority to accept gifts, bequests and devises of property, both real and personal, for the purpose of aiding or facilitating the work of NAGB, as provided under Section 421 of the Department of Education Organization Act, in accordance with Departmental procedures established in the June 4, 1991 memorandum to Executive Officers from the Director of the Budget Systems Division with respect to cash contributions and, with respect to noncash gifts, with the concurrence of the Office of the General Counsel;

(e) the authority to obtain the services of experts and consultants as provided under Section 402 of the Department of Education Organization Act;

(f) the authority to make, enter into, and perform contracts, leases, or other similar transactions as provided under Section 415 of the Department of Education Organization Act;

(g) the authority to review and approve for publication or public dissemination, articles, reports, brochures, speeches, books and other documents prepared by the Board and its staff.

(3) Except as provided in subparagraph (b), the Department shall provide administrative support services to the Board in accordance with applicable law, rules, regulations, policies and procedures. In providing administrative support services, the Department shall follow its own policies and procedures until the Board adopts alternative policies and procedures that are approved by the appropriate authorities.

(a) With respect to personnel matters, the Department shall provide advice and support services to the Board.

(b) With respect to contract matters, the Department shall provide advice to the Board, but the Board shall be responsible for the performance of contracting functions.

Lamar Alexander
Secretary of Education

March 4, 1992
Date

Richard W. Boyd
Chairman, National Assessment Governing Board

March 5, 1992
Date

BEST COPY AVAILABLE

National Assessment Governing Board's Response to Outside Recommendations

This appendix presents background information on the Governing Board and standard setting as well as (1) an overview of our recommendations and the Governing Board's responses to them, (2) an overview of the National Academy of Education's (NAE) recommendations and the Governing Board's response to them, and (3) some perspectives on the Governing Board's approach to standard setting in view of NAE's, our, and other experts' reviews.

Background

The National Assessment of Educational Progress' (NAEP) 1988 authorization assigned the Governing Board to identify appropriate achievement goals for each age and grade in each subject area tested under the national assessment. To carry out this responsibility, the Governing Board set achievement levels for NAEP as performance standards that define what students in grades 4, 8, and 12 should know and be able to do in each subject area. In pursuing this task, the Governing Board identified scores (called cut points) on the NAEP scale for basic, proficient, and advanced achievement levels for each grade and used these levels in reporting most NAEP results in 1990, 1992, and 1994. Using performance standards to report NAEP scores (criterion-referenced) rather than just reporting the distribution of test results (norm-referenced) was viewed by the Governing Board as a way to provide more useful and meaningful information to policymakers and the public.

In developing achievement levels based on performance standards, the Governing Board faced many policy, procedural, and technical issues. In addition, educational practice offered little guidance on the best approach for setting standards for broad-based assessments such as NAEP. After consulting the National Center for Education Statistics (NCES) and other testing experts, the Governing Board chose a well-known standard-setting method, an item-centered approach, but modified it in novel ways.[30]

Given the national significance of NAEP tests, the Governing Board's process for developing achievement levels was widely scrutinized and evaluated. NCES officials, research firms under contract either to the Governing Board or to the Department of Education, and other independent experts conducted reviews of the process. Two such reviews included our 1993 study focusing mainly on the Governing Board's 1990

[30]In item-centered approaches, panelists judge how students who have the capabilities needed to meet a given standard would perform on each test item. On the basis of the panelists' judgment, a cut point is identified as the threshold score for meeting the standard.

standard-setting process[31] and the NAE's 1993 study mainly of the 1992 process.[32] NAE's study was done at NCES' request.

Overview of Our Recommendations and the Governing Board's Responses

In 1993, we reviewed the Governing Board's approach for establishing the 1990 performance standards for NAEP's mathematics assessment. We analyzed the validity of issues raised by the Governing Board's approach, examined alternative approaches to setting performance standards, and reviewed the Governing Board's capacity to provide sound guidance to the Department of Education on technical issues. The Governing Board's 1990 standard-setting approach was procedurally flawed, resulting in interpretations of scores of which the validity was doubtful, according to our review. We recommended reopening the question of setting and interpreting performance standards. We also recommended reviewing the structures and procedures governing the assessment to ensure the technical soundness of the Governing Board's policies as well as its responsiveness to constituent interests.

We made a total of 10 recommendations to the Governing Board. In reviewing the status of these recommendations, we determined by October 1995 that the Governing Board had implemented 5 of the 10 recommendations but had not fully responded to the remaining 5. For example, the Governing Board had reexamined its approach to standard setting but chose to continue using the item-centered approach with some modification. The Governing Board also took some steps to strengthen its capacity to give sound policy advice to the Department by, for example, instituting a formal procedure to obtain NCES review of its proposed policies. The 10 recommendations and the Governing Board's responses as determined by us are summarized in table III.1 at the end of this appendix.

Overview of NAE's Recommendations and the Governing Board's Responses

On the basis of several studies it commissioned to review the Board's process for setting performance standards, NAE found that the Board's item-centered approach was fundamentally flawed for setting the 1992 performance standards and that the standards were set unreasonably high. NAE also concluded that the flawed performance standards would not enhance the ability to interpret NAEP scores and could harm NAEP's credibility. NAE called for the Governing Board to stop using an

[31]Educational Achievement Standards: NAGB's Approach Yields Misleading Interpretations (GAO/PEMD-93-12, June 23, 1993).

[32]Lorrie Shepard, Robert Glaser, Robert Linn, and George Bohrnstedt, Setting Performance Standards for Student Achievement, A Report of the National Academy of Education Panel on the Evaluation of the NAEP Trial State Assessment: An Evaluation of the 1992 Achievement Levels (Stanford, Calif.: Stanford University, 1993).

item-centered approach to setting performance standards and not report the 1992 NAEP results using these standards. NAE also determined that the 1992 NAEP test items, particularly at the advanced level, were not sufficiently congruent with emerging national content standards.[33] Consequently, NAE reported that the public may perceive that the expectations for developing students' thinking and mastery of challenging subject matter were low. Among its recommendations, NAE urged the Governing Board to wait until the availability of national content standards and then follow a more coherent process for developing performance standards in conjunction with content standards. NAE also urged the Governing Board to explore setting standards using several different approaches, none of which were item centered.

On the basis of its review of the NAE study and reviews by outside experts, the Governing Board chose to continue using an item-centered approach. The Governing Board has researched and experimented, however, with alternative approaches for setting the 1994 and 1998 standards, it has reported. It has also reported that it has widened its participant base in setting the performance standards to ensure a more thorough national consensus in setting the standards.

Perspectives on the Governing Board's Continued Use of an Item-Centered Approach to Standard Setting

Because the main issue underlying our and NAE's recommendations was the Governing Board's item-centered approach to standard setting, we consulted Governing Board officials and records to explain why the Board continued to use this approach—with some modifications. In general, Board officials said that they had considered our and NAE's recommendations, explored alternative standard-setting approaches, sought advice from many different experts, and made some modifications when their research warranted it but still concluded that the item-centered approach was the best available. The following examples provide some perspective on the Board's position:

[33]The movement to establish national content standards in the various disciplines began, as NAE describes, with the 1989 Education Summit when President Bush and the nation's governors identified six broad education goals. The creation of the goals led to initiatives to establish national content standards, including the Congress' creation of the National Council on Education Standards and Testing (NCEST). Several groups, such as the National Academy of Sciences/National Research Council, National Council of Teachers of English, and National Council of Teachers of Mathematics, have already developed or are developing content standards in mathematics, history, civics, geography, English, and other disciplines. NCEST distinguishes content standards from performance standards as follows: Content standards describe the knowledge, skills, and other understandings that schools should teach for students to attain high levels of competency in challenging subject matter; performance standards define various levels of competency in the challenging subject matter set out in the content standards.

- The Governing Board did not, as we recommended, withdraw its instructions to the Department for reporting the 1992 NAEP results because, after the 1990 effort and before beginning the 1992 process, Governing Board officials said they consulted with representatives from NCES and several national research firms and education associations that unanimously advised them that its standard-setting approach was the best known procedure.[34]
- As we reported in 1993, the Governing Board knew of some of the flaws in its 1990 procedures and acted to correct them for 1992. For example, instead of a group of ad hoc advisers, the Board secured an experienced contractor to set the 1992 standards who made improvements that strengthened the item-centered approach. These included careful attention to panelist selection, improved training, development of guiding definitions for each subject and grade before beginning the item judgment process, and review of the reliability of the judgment results. We concluded, however, that these changes did not address the fundamental problem of finding a test score that the Board could interpret validly using its definitions and descriptions.
- Regarding NAE's recommendation urging the Governing Board to altogether discontinue using its item-centered approach for setting standards, the Board contracted with two national testing experts to conduct an independent review of NAE's study. After evaluating NAE's study and the background studies, both experts concluded that the evidence did not warrant NAE's conclusions. State testing directors who reviewed NAE's study and heard the experts' rebuttal concluded that the Board should continue using its standard-setting approach, Board officials reported.
- The Governing Board and NCES jointly sponsored an October 1994 national conference of testing experts to discuss setting standards for large-scale assessments. No consensus emerged from this conference regarding an alternative method to be used in place of the item-centered approach.
- Before awarding the 1994 contract for developing performance standards, the Governing Board instructed all offerors to propose alternative designs and methodologies and to respond to issues raised by NAE and us, according to a Governing Board summary of its response. The Board reported that the contractor searched the literature and sought advice from many national and international consultants but found no better alternatives to the item-centered approach. The contractor also explored new approaches that called for judges to evaluate whole test booklets but reported that the judges chose to evaluate items individually.

[34]The Governing Board reports that representatives from the Educational Testing Service, the National Council of Teachers of Mathematics, American College Testing, and the Mathematics Sciences Education Board were included in this meeting.

- The contractor is experimenting with other approaches, including some recommended by NAE, to setting performance standards for the 1998 assessment, the Governing Board reported.

Table III.1: Our 1993 Recommendations and the National Assessment Governing Board's Actions Taken in Response

Our recommendation[a]	Governing Board action[b]
Because the current Governing Board's approach to setting standards has yielded unsupported interpretations of NAEP scores, we recommend that	
1. The Governing Board withdraw its instructions to NCES to publish 1992 NAEP results primarily in terms of achievement levels. The conventional approach to scoring interpretation should be retained until an alternative has been shown to be sound.	1. Action taken not fully responsive. The Board did not follow this recommendation, but it did change its methods and reporting practices to guard against misleading interpretations of the levels.
2. The Chairman of the Governing Board and NCES Commissioner develop a joint plan and schedule for a review of the Board's achievement levels approach, taking into account evaluations under way and providing for additional activities as needed. The NCES Commissioner should determine whether the Governing Board's approach will produce invalid interpretations of NAEP scores and should not be pursued or whether the approach is sufficiently promising that a specific plan for preparing for NCES prepublication review should be designed and implemented.	2. Action completed. A conference to examine standard-setting methods was held in 1994. Methods have been amended.
3. The Governing Board should withdraw its policy of applying the 1990 achievement levels approach to future NAEP tests and join with NCES in exploring alternatives for setting both content-based and overall performance standards regarding NAEP. This inquiry should examine issues of purpose, technical feasibility, cost, fairness, credibility, and usefulness.	3. Action taken not fully responsive. The Board did not withdraw its policy, but it did change its approach.
To strengthen the Governing Board's capacity to give sound policy direction, we recommend that the Governing Board	
4. Obtain NCES review of the technical strengths and weaknesses of proposed policies that implement the Governing Board's statutory responsibilities before making final decisions on such policies.	4. Action completed. The Board instituted a formal procedure to obtain NCES review.
5. Analyze the probable effect of proposed policies on NAEP's ability to present achievement fairly and accurately and to support both valid and reliable trend reporting.	5. Action taken not fully responsive. The Board's response stated that policies already in place provide for adequate monitoring and proposed no additional action.
6. Pilot test and thoroughly evaluate any new design or analysis procedure before it is fully implemented and results are reported.	6. Action completed. The 1992 and 1994 procedures were pilot tested.
7. Adopt standards of technical quality (to be applied internally) for publications issued under its own authority and also secure competent external technical review of such publications before authorizing their release.	7. Action taken not fully responsive. The Board reported it had no plan to issue publications that would require the adoption of technical standards but that if such plans arose, it would apply standards and reviews.

(continued)

Our recommendation[a]	Governing Board action[b]
8.The Chairman of the Governing Board should review actions taken regarding its statutory responsibility in the past 2 years, identify those whose technical consequences have not been sufficiently examined, and secure technical review as necessary to ensure that these actions will generate no unanticipated technical difficulties in the future.	8. Action completed. The Board has undertaken the review.
9. The Chairman of the Governing Board should review proposed policy to ensure that the Board prescribes policy ends, not technical details of implementation.	9. Action taken not fully responsive. The Board stated that its policy will remain in force but proposes no specific actions.
10. Regarding the Governing Board's membership, the Board should nominate for the testing and measurement positions people who are trained in designing and analyzing large-scale educational tests. The Board should also ensure that two or more of its elected officials, educators, and representatives of the general public have significant technical knowledge and experience.	10. Action completed. The Board reports that new appointees have been appropriately qualified.

[a]Our recommendations are from Educational Achievement Standards: NAGB's Approach Yields Misleading Interpretations (GAO/PEMD-93-12, June 23, 1993).

[b]We identified the Governing Board's actions by reviewing the status of our recommendations (made in report GAO/PEMD-93-12) as of Oct. 31, 1995.

Comments From the National Assessment Governing Board

THE NATION'S REPORT CARD naep

National Assessment Governing Board

National Assessment of Educational Progress

May 27, 1998

Ms. Carlotta C. Joyner, Director
Education and Employment Issues
U.S. General Accounting Office
Washington, D.C. 20548

Dear Ms. Joyner:

On behalf of the National Assessment Governing Board, I appreciate the opportunity to comment on the draft GAO report entitled, Student Testing: Issues Related to Voluntary National Mathematics and Reading Tests, which was transmitted on May 19, 1998.

Overall, the report is fair, balanced, and accurate. We appreciate the care and objectivity with which GAO explains a number of the complex issues and structures involved in the Governing Board's role in developing the proposed voluntary national tests.

We do have a few comments on specific details. These follow below.

Page 10: *"The Governing Board expects to use the same approach for setting performance standards for the voluntary national tests as it does for the NAEP assessments."*

This sentence may be misleading.

According to the House-Senate conference report on P.L. 105-78, the proposed national tests, designed for individual students, are required to "be based on the same content and performance standards as are used for NAEP," which is given to large groups of students on a representative-sample basis. The conference report directs that the proposed tests be "linked to NAEP to the maximum extent possible."

To carry out these directives, the contract and specifications approved by the Governing Board require that the proposed tests be designed to show whether individual students can meet NAEP standards for basic, proficient, or advanced performance. This will be done through statistical procedures to link various forms of the tests. No new standard-setting is anticipated. However, the Board may make some modifications or amplification in the written descriptions of achievement levels to improve the clarity and usefulness of the performance standards.

800 North Capitol Street, N.W.
Suite 825, Mailstop 7583
Washington, D.C. 20002-4233
Phone: (202) 357-6938
Fax: (202) 357-6945

Pages 33-35: *Test Specifications Contract*

The GAO should note that the test specifications adopted by the Governing Board on March 7, 1998, supersede those recommended by the expert panels convened under the MPR contract. It is these specifications, approved by the Board, that are being used to prepare national test questions and booklets. They are based closely on the assessment frameworks, developed through a national consensus process, that the Board previously adopted for NAEP. The Board action in March was taken upon recommendation of NAGB's two subject-matter committees and followed three months of deliberation, including public hearings and technical reviews.

Under the adopted specifications, both the reading and math tests would cover content and skills that are independent of any particular teaching method or approach. Credit would only be given for correct answers.

In the reading test, passages and questions would be similar to the grade four NAEP. All answers would require reference to the text presented. No questions will ask students about personal experiences, attitudes, or feelings.

In the mathematics test, the content includes computation, measurement, geometry, data analysis, and algebra, covered in the same proportions as the eighth grade NAEP. Computation will be assessed by itself and in solving problems. Items on which students may use calculators will comprise about one-third of the exam, the same as on NAEP.

There will be nothing fuzzy about these exams. Just as in NAEP, students will have to reason and explain how they reach their conclusions or results. In math they will have to compute. They will also have to solve problems. But no one will get credit just for trying.

On a smaller point, the table on page 11 should note that there is no federal requirement for participation in either NAEP or the voluntary national test proposal. In NAEP, however, once a state or district agrees to participate, the issue of whether students or schools must take part in the assessment is a matter for the state, district, or school to decide. In practice, who decides varies widely around the country.

Again, let me express appreciation for carefulness and clarity of the GAO report.

Sincerely,

Roy Truby

Roy Truby
Executive Director

Comments From the Department of Education

UNITED STATES DEPARTMENT OF EDUCATION
OFFICE OF THE ASSISTANT SECRETARY
FOR EDUCATIONAL RESEARCH AND IMPROVEMENT

THE ASSISTANT SECRETARY

June 1, 1998

Ms. Carlotta C. Joyner
Director, Education and Employment Issues
General Accounting Office
Washington, DC 20548

Dear Ms. Joyner:

Thank you for the opportunity to provide comments for the Department of Education on the GAO Draft Report, "Student Testing: Issues Related to Voluntary National Mathematics and Reading Tests."

The voluntary national tests in fourth-grade reading and eighth-grade mathematics will be important tools for helping teachers and parents improve teaching and learning in our schools. The three key advantages of voluntary national tests are simple, but powerful. The tests will (1) measure student performance against widely agreed upon standards of what students should know and be able to do; (2) focus attention on improving achievement in the basic subjects of reading and mathematics; and (3) provide states, school districts, and schools with detailed information to evaluate the effectiveness of their educational programs.

In August 1997, Secretary Riley proposed that the authority for assessing the development of the voluntary national tests be placed with the National Assessment Governing Board (NAGB). We are pleased that the Department's Appropriations Act for fiscal year 1998 gave NAGB the authority to begin test development. As your report demonstrates, both the Department and NAGB have fully honored the requirements of that Act, and NAGB has assumed responsibilities for test development. NAGB has made good progress and should continue to move forward.

Since President Clinton announced the voluntary national test initiative in his 1997 State of the Union address, there has been local demand for the tests. Seven states and sixteen school districts, representing more than 20 percent of fourth- and eighth-graders, have signed up to use the tests. We should give states and local districts the chance to use tools, like the voluntary national tests, to improve local schools and increase our students' mastery of reading and mathematics.

Sincerely,

Ricky T. Takai
Acting

555 NEW JERSEY AVE., NW, WASHINGTON, D.C. 20208

BEST COPY AVAILABLE

GAO Contacts and Staff Acknowledgments

GAO Contacts

Eleanor L. Johnson, Assistant Director, (202) 512-7209
Barbara A. Billinghurst, Evaluator-in-Charge (206) 287-4867

Staff Acknowledgments

In addition to those named above, the following individuals contributed to the report: Peter J. Bylsma helped plan the job, collect information, and prepare an early draft of the letter; Stanley G. Stenersen helped organize the report topics and write report sections; Julie A. Rachiele searched databases for textbook authors; and Mary W. Reich and Dayna K. Shah provided legal advice.

Ordering Information

The first copy of each GAO report and testimony is free. Additional copies are $2 each. Orders should be sent to the following address, accompanied by a check or money order made out to the Superintendent of Documents, when necessary. VISA and MasterCard credit cards are accepted, also. Orders for 100 or more copies to be mailed to a single address are discounted 25 percent.

Orders by mail:

U.S. General Accounting Office
P.O. Box 37050
Washington, DC 20013

or visit:

Room 1100
700 4th St. NW (corner of 4th and G Sts. NW)
U.S. General Accounting Office
Washington, DC

Orders may also be placed by calling (202) 512-6000 or by using fax number (202) 512-6061, or TDD (202) 512-2537.

Each day, GAO issues a list of newly available reports and testimony. To receive facsimile copies of the daily list or any list from the past 30 days, please call (202) 512-6000 using a touch-tone phone. A recorded menu will provide information on how to obtain these lists.

For information on how to access GAO reports on the INTERNET, send an e-mail message with "info" in the body to:

info@www.gao.gov

or visit GAO's World Wide Web Home Page at:

http://www.gao.gov

PRINTED ON RECYCLED PAPER

BEST COPY AVAILABLE

United States
General Accounting Office
Washington, D.C. 20548-0001

Bulk Rate
Postage & Fees Paid
GAO
Permit No. G100

Official Business
Penalty for Private Use $300

Address Correction Requested

SHIP TO:
CHRISTINA ROWSOME
ERIC PROCESSING & REF FACILITY
2ND FL
1100 WEST ST
LAUREL MD 207073587

Request # 4053896 70 3

07/06/98 / 16:35 m
07/06/98 / 21:15 £

4053896011p205
Item Document Number
1 HEHS-98-163

BEST COPY AVAILABLE